#### **The Front-end ASIC for the ATLAS Pixel Detector**

# K. Einsweiler, LBNL on behalf of ATLAS Pixel Collaboration

Overview of FE specifications and design

**History of ATLAS Pixel FE ASIC** 

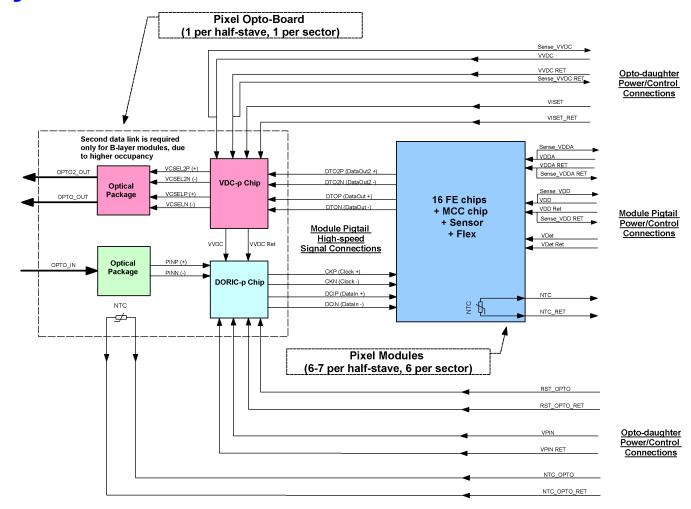
The first 0.25 $\mu$  generation of the FE ASIC, FE-I1

Performance results using FE-I1

**Second generation 0.25**μ **FE ASIC**, **FE-I2** 

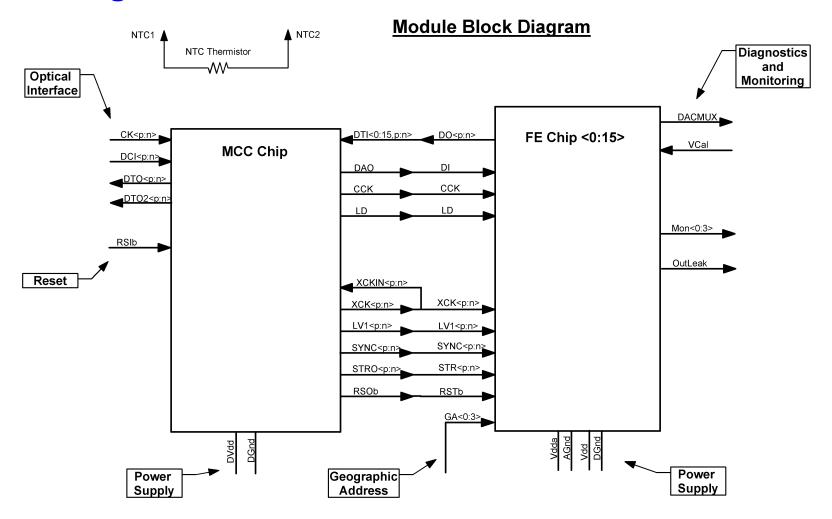
## **System Design of Pixel Module**

#### **Overall system architecture:**



- Optical package and DORIC/VDC mount on separate opto-card, up to 1m away.
- Module itself uses two LV supplies (analog and digital) and one HV bias supply.
- Communication between module and opto-card uses 3mA LVDS I/O

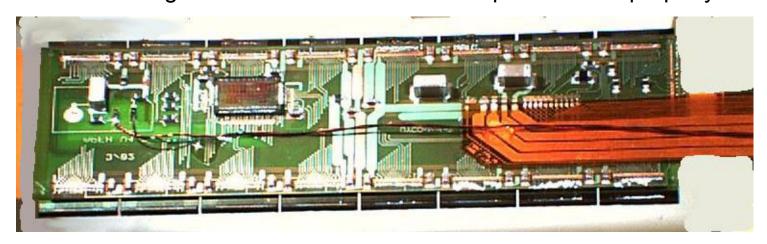
## **Block diagram of module itself:**



- Two chip design, including a single controller and event-builder chip (MCC), and 16 front-end chips bumped to a single silicon sensor substrate.
- •Flex hybrid is used to provide interconnections above.
- Module has 46,080 channels, 10 cm<sup>2</sup> active area, and dissipates up to 5W.

#### **Features:**

- Basic interface to the outside uses a 3-wire protocol (SerialIn, SerialOut, XCK).
- •Basic interconnections between FE and MCC use bussed signals. Slow control uses full-swing CMOS. Fast signals use low-swing differential LVDS signaling.
- •To provide enhanced speed and robust module design, the serial output lines are connected from the FE to the MCC in a star topology (16 parallel inputs on MCC).
- •There are no analog signals between MCC and FE. All FE chips have internal current references and DACs to control analog operating points and calibration.
- •Architecture is "data-push" style: each crossing for which LV1 accept is present causes all FE chips to autonomously transmit back hit information for the given crossing. LV1 signal may remain set for many contiguous crossings to allow readout of longer time intervals (up to 16). MCC merges such events together.
- •Synchronization signal available to ensure FE chips label LV1 properly.



# **Basic FE Chip Geometry**

- Agreement on pixel size was struck in Sept 96, in order to allow compatible, parallel detector and electronics development.
- •The geometry adopted was  $50\mu$  x  $400\mu$  for the pixel size, with pixels arranged in 18 columns of 160 pixels per column.
- •The geometry was mirrored between columns, so that inputs for pixels in column 0 and 17 are on the outside. All other inputs are paired. This gives us 9 column pairs, with a common digital readout in the center, and analog cells on the sides.
- •The input pad geometry in the inner column pairs is then a double row of  $50\mu$  pitch pads. The metal pad is specified to be  $20\mu$  octagon, with a  $12\mu$  opening in the passivation for the bump-bonding.
- •The cut die size must not extend beyond  $100\mu$  from the edge of the active area on three sides of the die. Hence, nothing outside of the pixel circuitry is allowed on three sides of the chip, to allow module construction.
- •The bottom of the chip (all peripheral logic and I/O pads) is allowed 2800 $\mu$ , making the total active die region 7.2mm x 10.8mm.
- •An I/O pad structure of 48 pads, each consisting of a  $100\mu$  x  $200\mu$  wire-bond pad, and a group of 4 bump-bond pads for MCM-D applications, was frozen.
- •For final modules, only the central 30 bond pads are available for connections due to mechanical envelopes. Other 18 pads are available for diagnostics.

# **Brief History: FE-B, FE-D, and FE-H**

- •Rad-soft prototyping delivered functional chips in 98 (FE-B, FE-A/C). FE-B (HP 0.8μ) demonstrated all basic ATLAS pixel performance goals in lab and testbeam.
- •Submitted FE-D1 run, containing FE-D1 front-end chip, DORIC and VDC chips, and prototype MCC-D0 chip (plus test chips). Submission went out in July 99.
- •FE-D1 suffered from minor design errors, and very poor yield in two circuit blocks. After considerable investigation, the low yield was related to technology problems (low rate of very leaky NMOS). Foundry never succeeded in isolating the problem, but proposed a series of special corner runs.
- •Submitted FE-D2 run in Aug 00, with two versions of FE-D2. In one version, all design errors were fixed, but basic design (including dynamic logic blocks which suffered low yield) was left unchanged. Second version replaced low-yield blocks with static versions, and removed other circuitry (trim DACs) to make room. Run included full MCC-D2 (100mm²) and new DORIC and VDC chips as well.
- •Corner runs gave no new information on yield/technology problems. Yield on static chip was better, but still unacceptable. Work with this vendor was terminated.
- •Began work on FE-H in Dec 99. Chip was almost ready to submit when we received notification of massive cost increases from Honeywell. With wafer cost of 20-30K\$, effort was abandoned before actually building a complete pixel chip.
- •The failure of both traditional rad-hard vendors left us with 0.25μ approach, based on commercial process and rad-tolerant layout. Major effort started in Sept 2000.

# Summary of Key Requirements for FE Chip

- •Achieve in-time threshold of about 5Ke. In practice this involves operating with a 3Ke threshold, and requiring a 2Ke overdrive for in-time hit association. Imagine XCK phase is adusted so largest charges (100Ke) arrive just after leading edge. Definition of overdrive spec is that charge 2Ke above nominal threshold must arrive no more than 20ns later to be associated with the same crossing.
- •In order to operate module at 3Ke threshold with very low noise occupancy, must carefully control threshold dispersion and noise. Should have quadrature sum less than about 500e. Typically, this means threshold dispersion less than about 300e and noise less than about 400e.
- •Tolerate leakage current of up to about 50nA per pixel, achieve noise occupancy of less than 10<sup>-6</sup> hits/pixel/crossing, cross-talk to neighbors less than about 5-10%.
- •Double pulse resolution of about 0.5μs for innermost layer and 2μs for outer layer.
- •Associate all hits with unique 40MHz beam crossing, support L1 latency for all hits of at least 3.2μs, and triggered readout rate of up to 100KHz with negligible deadtime or hit loss.
- Make charge measurement for each hit of modest quality (4-5 bits).
- •Survive delivered dose of 50MRad and 10<sup>15</sup> 1MeV neutron equivalent.
- Provide ability to perform in situ testing and calibration of all functions and monitoring of all specifications to make sure performance does not degrade.

#### Feature List for FE-I1

# Design is logical evolution from FE-D and FE-H designs. Analog Front-end (designed for VDDA=1.6V operation):

- •The FE uses a DC-feedback preamp design which provides excellent leakage current tolerance, close to constant-current return to baseline for TOT, and very stable operation with different shaping times.
- •It is followed by a differential second amplifier stage, DC-coupled to the preamp. The reference level (VReplica) is generated in the feedback block, and should match the DC offset of the preamp with no input. The threshold control is applied using two currents to modify the offsets on the inputs to the second amplifier stage, allowing a large range for threshold control.
- The two-stage amplifier is followed by a differential discriminator which provides the digital output sent to the control logic.
- •The control logic provides a 5-bit threshold trim capability in each pixel, plus a 5-bit feedback current trim capability for tuning the TOT response. There are four control bits, including Kill (shut down preamp), Mask (block entry of hit into readout logic), HitBus (enable output to global FastOR) and Select (enable injection of charge for testing). The HitBus bit also controls the summing of a current proportional to the feedback plus leakage current in the preamp, allowing monitoring of the feedback current, and of the leakage current from the sensor.

- •A global FastOR net is created using all pixels enabled for this type of readout, and provides a self-trigger and diagnostic capability.
- •All critical bias currents and voltages on the chip are controlled by internal DACs. There are 12 8-bit DACs for the analog front-end, and an additional DAC for the charge injection. The current DACs are referenced to an internal CMOS current reference, and the DAC values are loaded from the Global Register, and controlled via the Command Register.
- •Two injection capacitors integrated into input pad. Clo has value of 5fF (actual 4.6fF), and Chi+Clo is about 40fF. VCal range of 0.8V gives 25Ke/200Ke ranges.

# Digital Readout (designed for VDD=2.0V operation):

- •It uses an 8-bit Grey-coded 40 MHz differential "timestamp" bus as a timing reference throughout the active matrix. All pixels measure their leading and trailing edge timing by asynchronously latching this reference in RAMs.
- Hits (address plus LE/TE timing) are transferred from the pixels as soon as the trailing edge occurs, using a shared bus structure in the pixel column pair. This bus operates at transfer rates up to 20 MHz in order to meet our requirements.
  Differential signal transmission and sense amplifiers are used to achieve this.
- •Significant buffering is provided in the end of column region for hit storage during the L1 latency (up to 6.4µs in this chip). Sixty four buffers are available for each column pair (one for each five pixels). The coincidence with the L1 trigger is performed in this buffer. Hits from rejected crossings are immediately cleared.

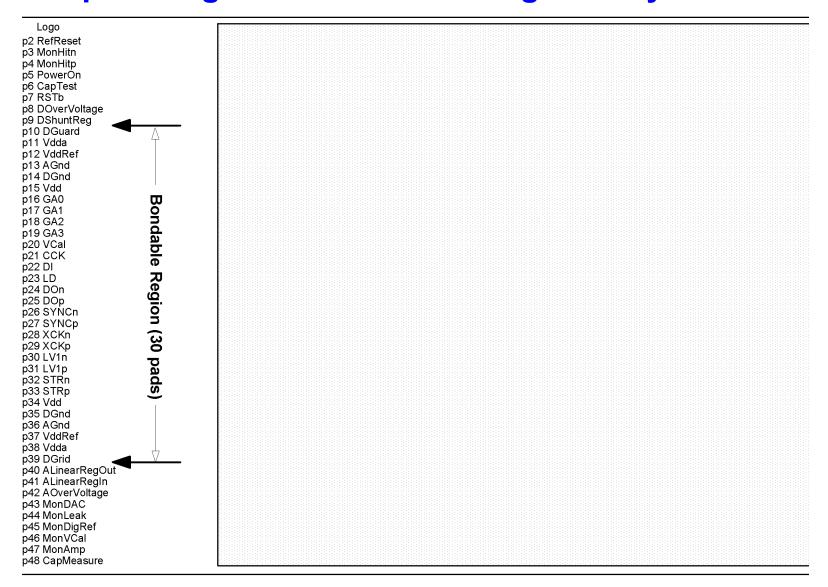
•A readout sequencer stores information on up to 16 events pending readout. As soon as the output serial link is empty, transmission of a new pending event begins. Essentially, sending a L1 trigger corresponds to making a request for the all hits associated with the corresponding beam crossing, which are then pushed off the FE chip to the MCC.

#### **Control Logic:**

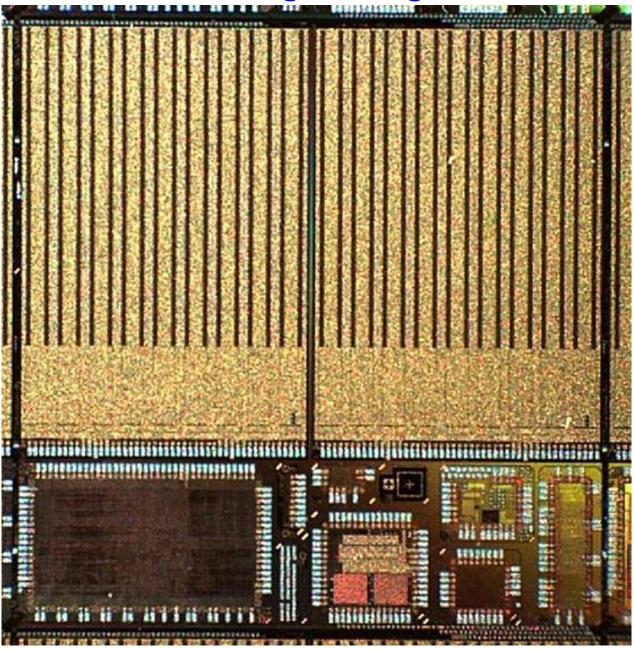
- •Global control of the chip is implemented using a simple command protocol. A Load signal controls whether input bits are interpreted as address and control, or as data. There is a 20-bit Command Register. Individual bits in this register implement specific commands (e.g. ClockGlobal, WriteGlobal, ReadGlobal).
- •A Global Register, consisting of 202 bits, controls Latency, DAC values, enabled columns, clock speeds, and several other parameters. This register is implemented as a shift register and a shadow latch with full readback capability. The shadow latch is SEU-tolerant since it contains critical configuration information.
- •A Pixel Register which snakes through the active array provides access to the 14 control bits in the pixel (Select, Mask, HitBus, Kill, FDAC<0:4>, TDAC<0:4>). Readback capability is supported by transferring FF information back into the long shift register for readout. The 14 latches in each pixel are SEU-tolerant.
- Each chip on a module is geographically addressed, and its identity is controlled by external wire-bonds to avoid confusion. A broadcast mode is also supported.

#### **FE-I1 Pinout and Geometry**

#### Sketch of pin assignments and overall geometry of die:



## **Reticle for FE-I1 engineering run:**



Reticle includes two different FE-I1 chips. Design has 2.5M transistors.

One is FE-I1A, with Cfb=10fF, the second is FE-I1B, with Cfb=5fF.

Reticle also includes MCC-I1, Analog Test Chip, LVDS Buffer chip, DORIC and VDC optochips, and several other small test chips.

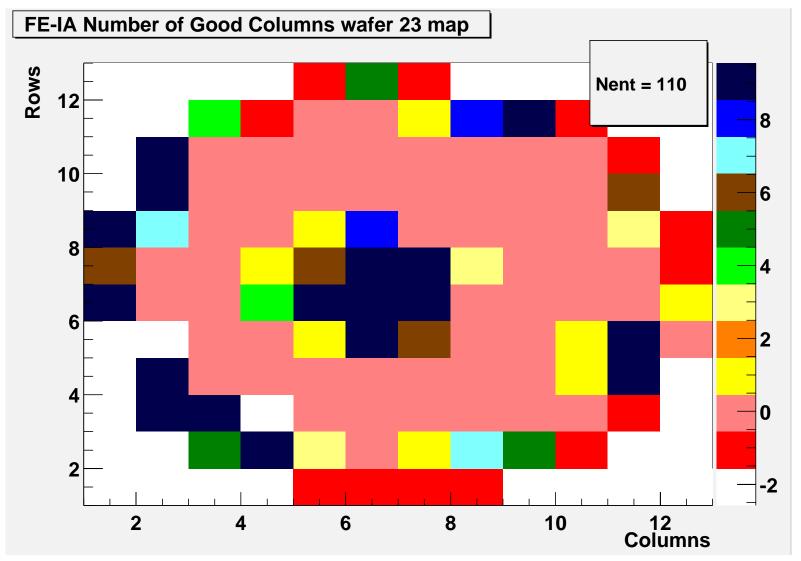
There are 112 potentially good reticles on a wafer.

#### **Initial FE-I1 Measurements**

## First wafers from 12-wafer Engineering Run arrived in Jan 02:

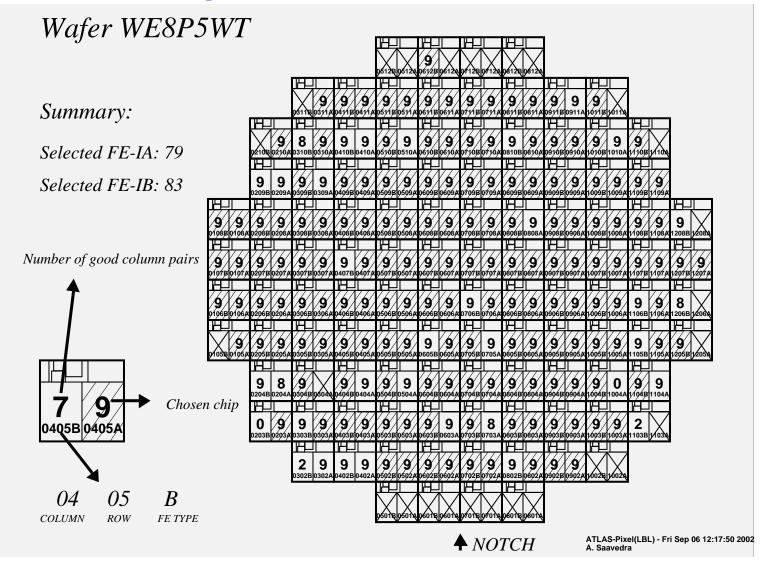
- •All blocks worked roughly as expected. Remarkable success for 2.5M transistor chip submitted in new process!
- •All performance features, even for new analog front-end, have performed close to expectations. Even threshold dispersion and timewalk, studied by large HSPICE simulations, agree reasonably well with the simulations.
- •However, it was quickly realized that there was a serious yield problem. Yield to pass simple selection criteria (analog/digital currents OK, all registers working, and basic digital inject test working) was only about 15%. Even more striking, the good chips were all confined to a small area in the core of the wafer, or along the extreme edges. Finally, chips that passed basic register tests (few percent of transistors tested), would usually be perfect for full digital and analog tests, so defect density was not an issue.
- •Extensive investigations of failure modes have been made, and fault analysis was performed by the foundry. Four additional wafers, initially held at back-end processing, were sent for evaluation. They showed very low yield (3%), and very basic failure modes consistent with metallization problems (mostly supply shorts).
- •Example of wafer map of first wafer probed shows typical pattern seen in first 12 wafers. We find very little variation within a lot or wafer group, but very large variations in yield between groups!

# Wafer Map for SESB23T (good column pairs for A chips):



•Chips with no data appear White, bad Global Registers are Red, and othe colors represent the Pixel Register test results. There are 18 (3) chips with working Global Registers and 9 (8) column pairs working in the Pixel Register.

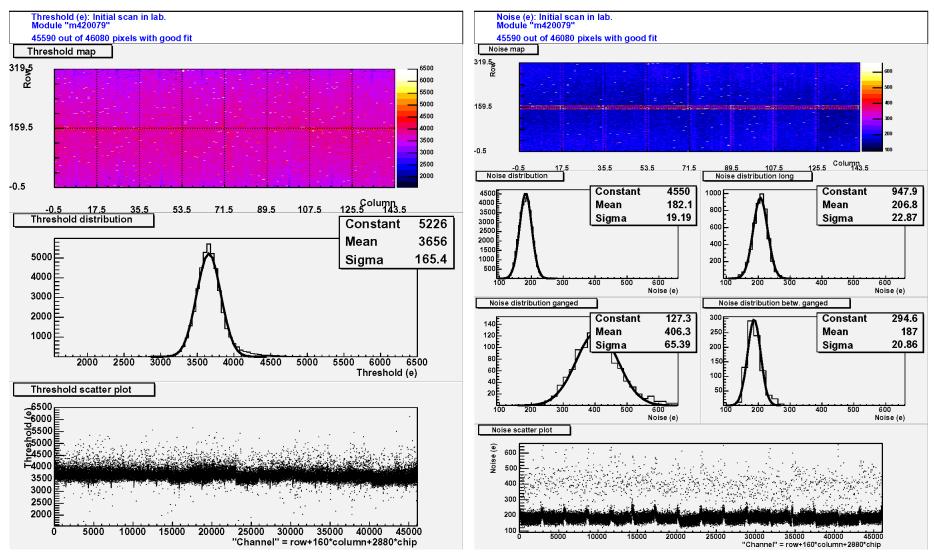
## Typical wafer from good replacement lot:



•Map shows chips passing complete cuts (supply surrents, registers, DACs, perfect digital pixels and EOC buffers). This has a yield of 73% for this wafer. With 24 wafers of this type, about 60 pre-production modules built and extensively studied.

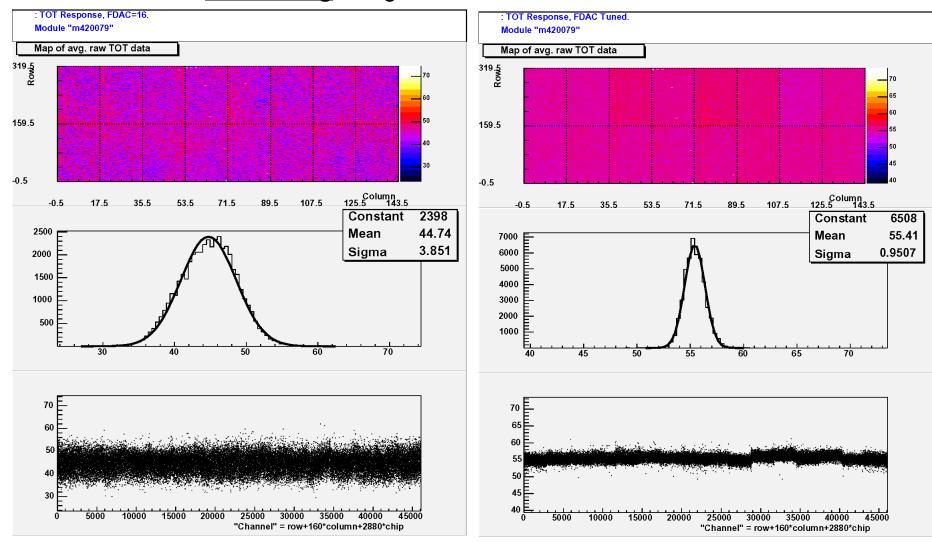
#### **Performance Measurements for FE-I1**

•Example of threshold and noise performance for module with IZM solder bumps:



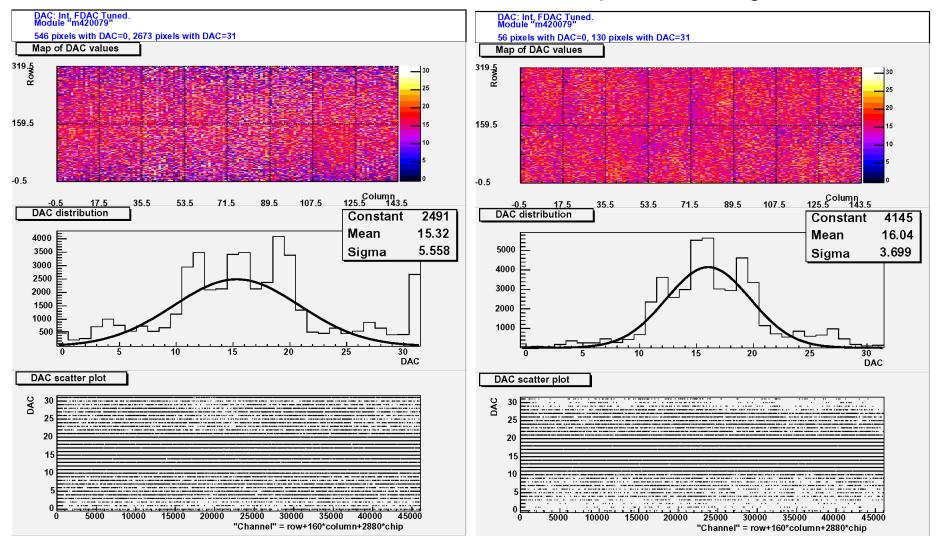
•Left plots show threshold maps over module, right plots show noise performance for different pixel types and maps. Typical dispersion about 150e, typical noise 180e.

#### Demonstrate <u>TOT tuning</u> using feedback DACs:



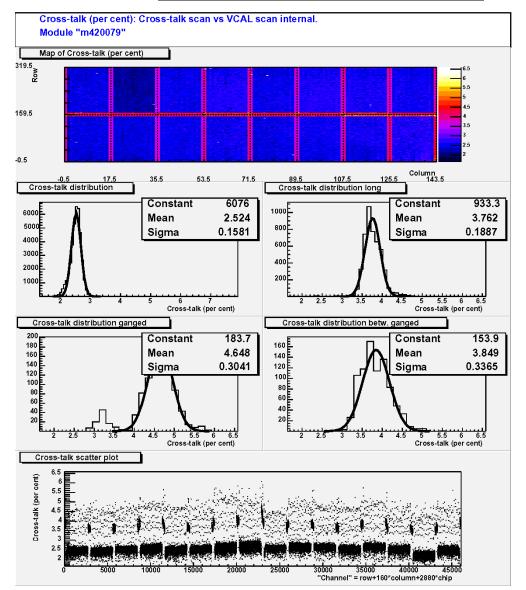
- Inject 20Ke signal in each chip (corrected for Clo and VCal), plot mean TOT. Adjusting 5-bit FDACs gives factor 5 reduction in TOT dispersion from 10% to 2%.
- Main issue is dealing with changes in TOT response with irradiation. TOT for a given charge increases by about a factor 2 with full irradiation dose.

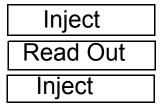
#### <u>Distribution of 5-bit TDAC and FDAC values</u> for previous tuning:



- •Plot on left is for TDACs. Non-Gaussian shape due to DAC non-linearities. Mixture of systematic and random effects are seen. Optimal tuning not achieved.
- •Plot on right is for FDACs. Same non-Gaussian distribution seen, but behavior is more random. No pile-up at ends of the distribution.

#### Measure <u>cross-talk between channels</u> in bump-bonded module:





Enable central pixel for readout, but disable charge injection.

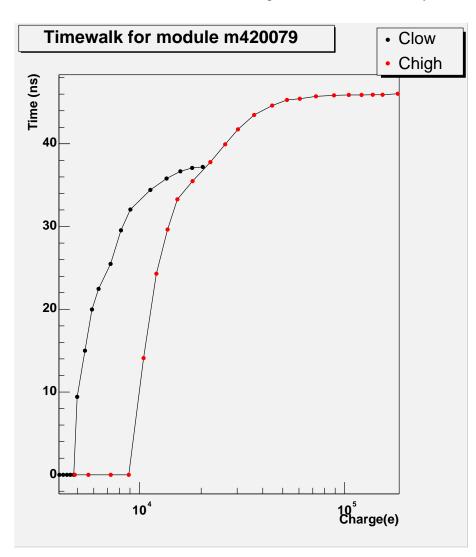
Disable readout for pixels n+1 and n-1, but enable injection.

Inject very large charges using Chi injection range.

Take ratio of threshold (n) divided by Q(n-1)+Q(n+1), where the sum of the Q is total injected charge required to fire the central pixel.

Want several MIP input not to cause neighbor pixels to fire.

 Typical cross-talk values for normal pixels are 2-3%, and special pixels (long, ganged, and long-ganged) pixels have 4-5% cross-talk, or factor 20 rejection. Measure <u>timewalk performance</u> (time slewing vs charge) in complete module:



Perform 2D scan: for given charge, scan delay time of charge injection and observe when hit is "in-time" with trigger. Small strobe delay corresponds to larger propagation delay (trigger time fixed).

MCC-I includes programmable 6-bit delay block with step size of about 1ns for this.

Use Chi scale to define large-charge asymptotic response. Use Clo to approach threshold. Note there is not good overlap in mid-range.

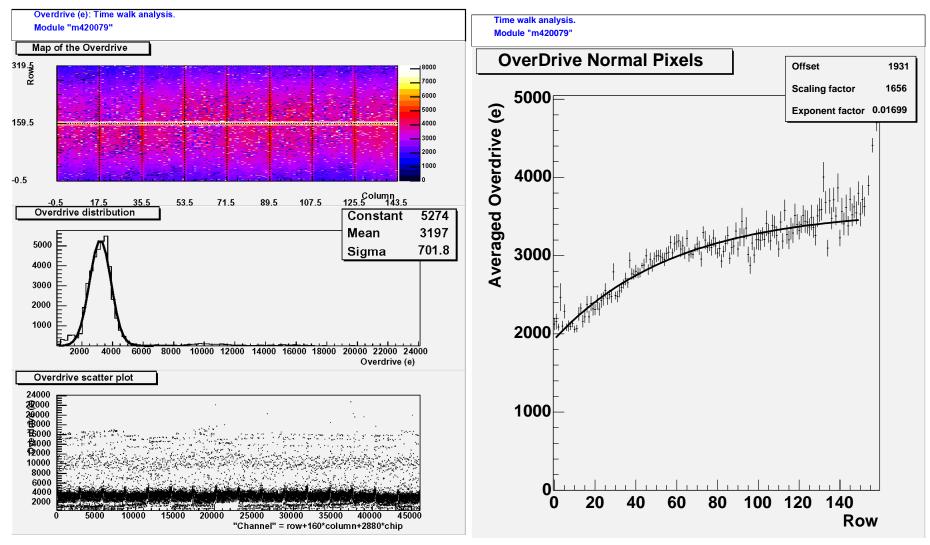
Allocate budget for FE chip of 20ns relative to large charge (typically 100Ke).

Calculate smallest charge which arrives within 20ns of large charge.

Subtract threshold to define "overdrive".

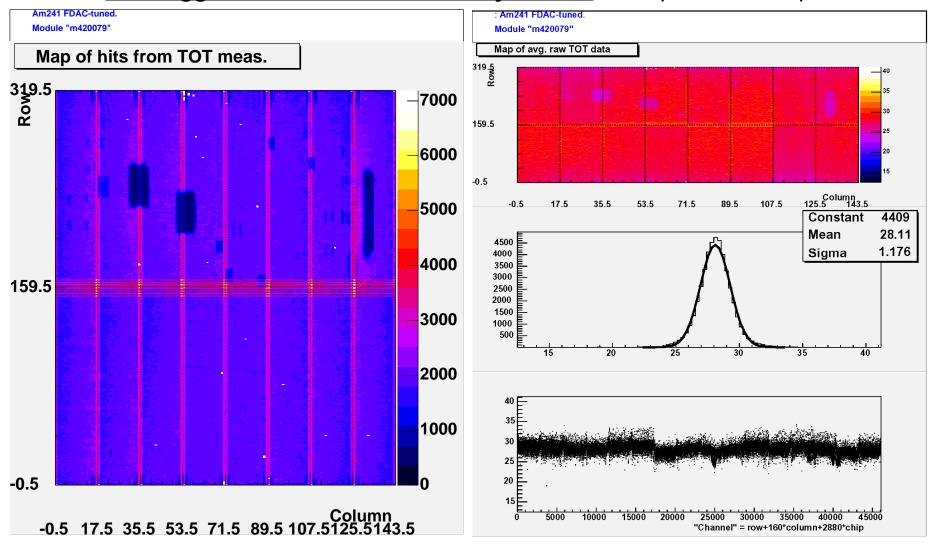
- •Requires a full 2D scan for each pixel, and with both different injection capacitor ranges. Curves above show "average" response for a module.
- Injecting large charge and looking at absolute time shows sigma of less than 1ns.

#### •Results of **timewalk measurement** (overdrive) for complete module:



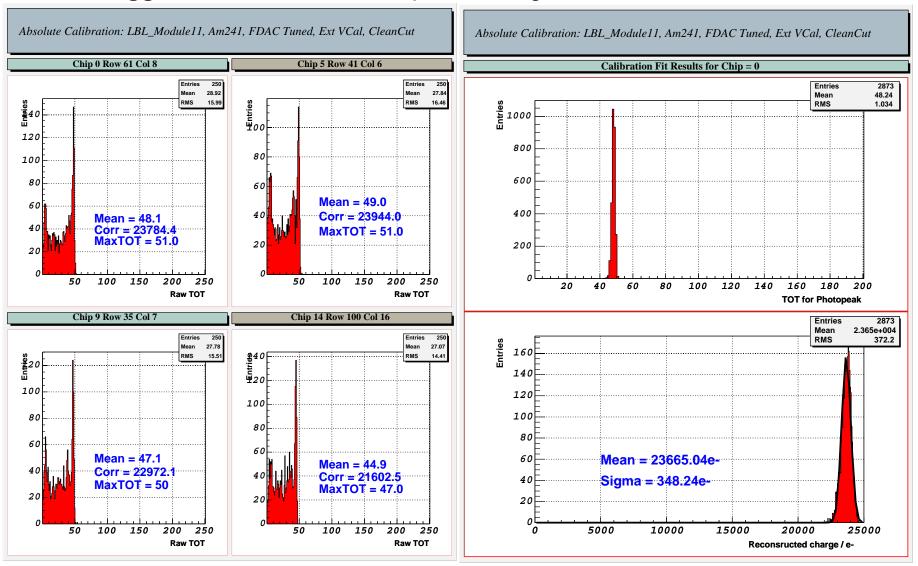
•Observe systematic problem of slower response at higher row numbers and for special pixels. Profile plot on right averages over all columns in module. Problem arises because bias current for input transistor depends on row number due to voltage drops on AGND net that are amplified by bias mirrors in weak inversion.

#### •Use **self-trigger mode with Am241 X-ray source** to map module response:



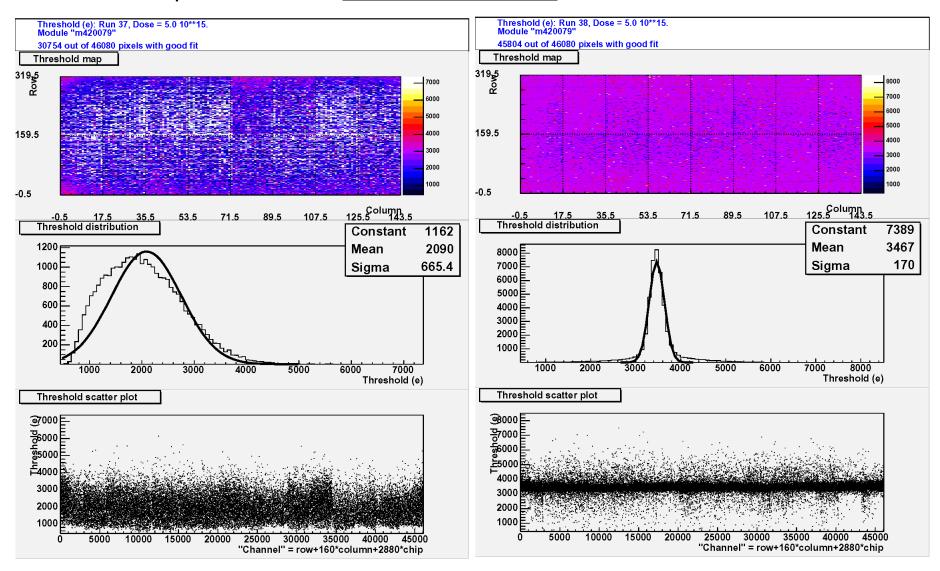
- •Left shows hit map for 5M trigger concurrent scan with 100mC source. Three large dark regions are LV and HV decoupling. Total of 35 dead or masked pixels here.
- •Right plots show average TOT response for each pixel. FDACs have been used to achieve uniform response. More interesting plot fits peaks in charge for each pixel.

#### Self-trigger Am241 Results with peak-finding and calibration:



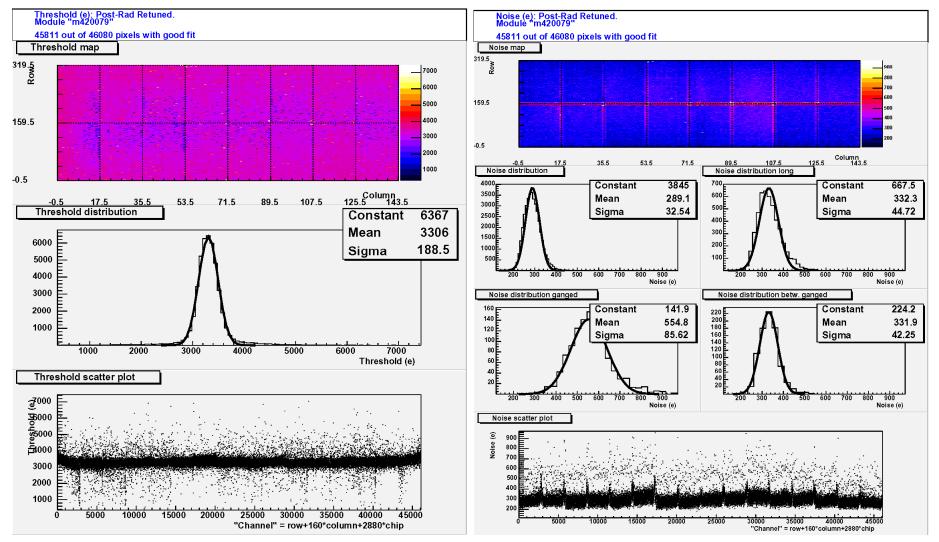
- Left plots show typical raw spectra for individual pixels.
- •Right plots show TOT for peak and calibrated charge for one chip. Absolute calibration and inter-chip calibration not yet fully understood.

#### Threshold performance of <u>irradiated modules</u>:



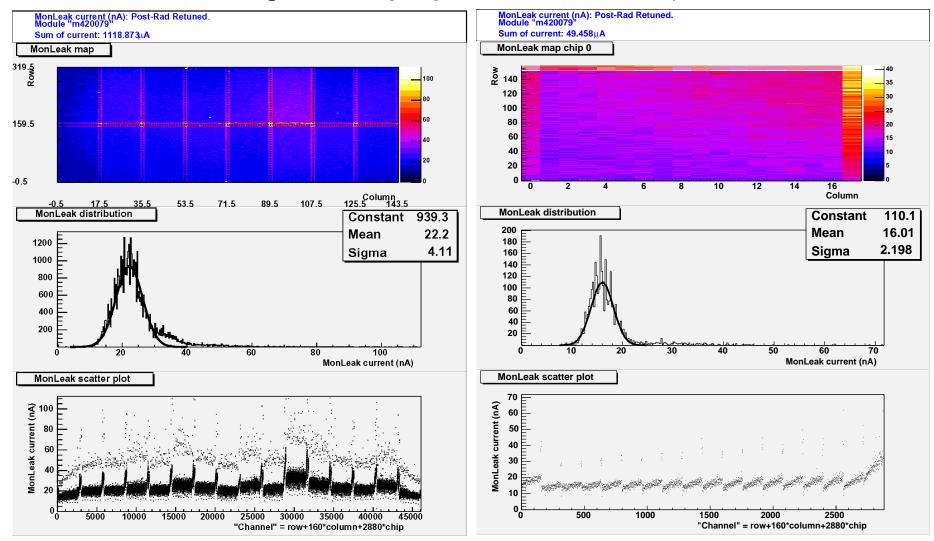
- •Thresholds in FE-I re-disperse under irradiation. After about 2MRad, initial tune has dispersion of about 300e and needs retuning.
- •Show in situ retune after 15MRad (left is before, right is after tune, not optimized).

#### • Threshold/noise performance of module irradiated to average dose of 30MRad:



- Dose was very inhomogeneous (factor 2 variation over surface of each chip), with maximum dose above 50MRad.
- •Retuned operation at -7C gives decent threshold and good noise performance. Tails towards low threshold are too large, and further optimization needed.

#### Can monitor leakage current per pixel in irradiated complete modules:



- •Plot on the left is for complete module, showing average current of 22nA/pixel. Plot on right is for chip #0 only, showing variation from about 10nA in lower left to 20nA in upper right.
- Proven to be a powerful tool in diagnosing sensor issues and monitoring doses.

# **Power Management**

- •Critical problem for large scale experiments like ATLAS. Power supplies located 140m from detector, with roughly 80% of voltage drop outside chips.
- •Use rad-hard regulators from ST at 12m from detector to compensate for large voltage drop on low mass services (2V round-trip), and provide correct voltage directly to flex hybrid of module. In limited system tests (6-7 modules) operating with full services prototype, this works. However, significant services burden!
- •Concern over transient problems in long services caused us to implement overvoltage protection circuits in FE-I1, using diode reference and large FET clamp with low resistance. Use of clamps has so far not proven to be necessary (have not killed any electronics during system testing).
- •Have also implemented power management inside individual chips for FE-I2. This includes a digital regulator that powers up to 1.8V, but can be trimmed up to 2.5V. There is a separate analog regulator that is initially inhibited, powers up at 1.5V, and can be trimmed up to 2.0V. Both regulators have low-dropout (about 100mV for 100mA), so only marginally increases the total power consumption. The regulators are always connected, but are "floating" if their inputs are not powered up (normal condition). Extensive simulations suggest this should be OK. Finally, in order to cope with steady state operation with up to 4V on input to the regulators, special diode protections implemented.
- •Approach may not be used in first generation detetector, but will certainly be needed for migrating to deeper sub-micron (0.13μ).

# Improvements for next generation (FE-I2)

•After significant experience with FE-I1, began an upgrade program in Summer 2002. Goal was to produce production quality chip for ATLAS. Difficult with only the second iteration in a new process.

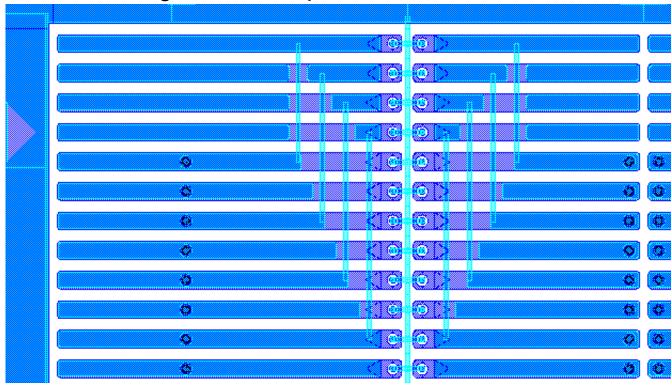
# Serious issues for FE-I2 (many minor bug fixes):

- Threshold dispersion and "re-dispersion": The initial dispersion for "high gain" version of FE-I1 was about 900-1000e at 3000e threshold. Can typically tune a given assembly to a sigma of about 100e for a given set of conditions. However, changing the temperature from +20C to -10C re-disperses the thresholds to about 250e sigma. Similarly, a total dose of about 1MRad re-disperses the thresholds to about 300e sigma. Measurements have shown that this re-dispersion does not saturate at high total dose. Finally, small changes in the global threshold adjustment also cause re-dispersion. Goal was to improve this performance.
- •Optimization of transistor sizes in preamp and second stage were predicted to decrease dispersion from 900e to 600e with essentially no impact on dynamic performance. Also, a Global threshold DAC was implemented in each pixel.
- •**Bias distribution:** Significant top-bottom variations are seen in the timewalk performance of FE-I1. These arise from internal voltage drops on AGND, which in turn modify the Vgs used to distribute large bias currents like IP (the preamplifier input transistor bias, typically 8µA). All mirror transistors in our design are in weak inversion (sub-threshold), so mirrored current is sensitive to small changes in Vgs.

**Threshold control:** Very compact local DACs used for 5-bit threshold adjust had poor linearity (not monotonic for certain bit combinations). In addition, the bias distribution issues mentioned above cause significant top/bottom variations. To optimize predictability of threshold tuning process, improved the quality of the local DAC. New threshold scheme implemented with differential, high quality DAC.

- Also implemented an "auto-tune" circuit to allow much faster adjustment of threshold trim DACs using up/down counter in each pixel.
- •**SEU-tolerance:** All configuration data (Command Register, Global Register, and Pixel Register) is stored in SEU-tolerant latches (40,547 per chip!) in FE-I1. The design used two sets of cross-coupled inverters to prevent upset if a single node was modified. Initial measurements of upset rates of our SEU-tolerant latches at the 55MeV Cyclotron showed low cross-sections. Measurements at the CERN PS (20GeV protons) showed more than an order of magnitude higher upset rates. This would lead to somewhat unreliable operation at the LHC, despite proposed global "periodic reset" every few hundred seconds in ATLAS TDAQ. Believe that problem was largely compact layout not latch design.
- •SEU-tolerance of FE-I2 improved by using optimized "bit-pair" layout of the latches with careful placement of all critical nodes (latch size about 3 times larger). Also have implemented a triple-redundancy scheme in the Command and Global latches, with redundant write and reset control. This should result in very stable configuration data even during operation of B-layer at design luminosity. Also add "Hit Parity" tag to each hit to look for single-bit errors during data handling.

- •SEU effects in data paths (dynamic) are much harder to evaluate or measure. All state machines designed to use no hidden states, and individual bit flips usually have localized effects (corruption/loss of individual hits). Have implemented SEU-tolerant DFF based on redundant inverter pairs and voting logic (similar to DICE).
- <u>Special pixels:</u> ATLAS pixel sensor contains four types of pixels, in order to provide 100% coverage in multi-chip module:



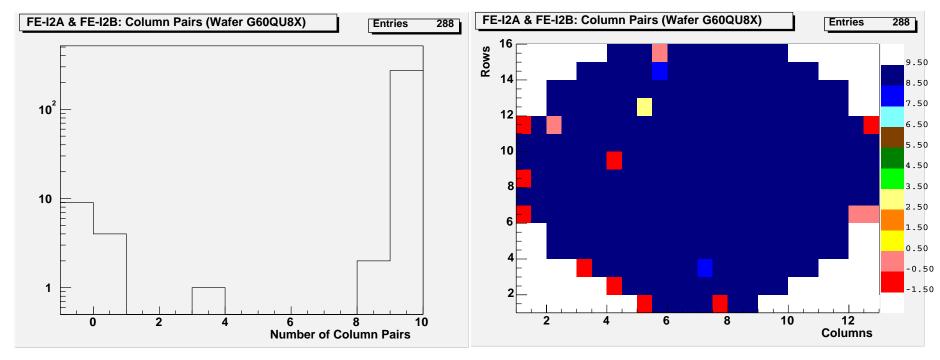
- •Pixels at the two edges are  $600\mu$  instead of  $400\mu$ . Pixels at the top are ganged.
- •Capacitive loads for ganged pixels, as well as normal pixels overlapped by ganging traces, are much higher, producing poor timing performance. Use modified frontends with 2\*IP and 4\*IP, in order to provide acceptable timewalk for all pixels.

#### **Status of FE-I2 Submission**

- •Final chip has 3.5M transistors (significant increase in auto-tune circuit). Design was migrated to 6-metal process to deal with extra connectivity requirements.
- Fabricated in IBM Burlington foundry instead of ALTIS foundry. First wafers received in mid-May.
- Testing revealed that almost everything worked perfectly. Major problem is timing errors in large place and route block at the bottom of the chip. This affects the global control of the chip. It is possible to operate in the lab at reduced VDD, but this is not appropriate for use in production. Major issue was that during SEU-hardening, size of this block increased by roughly a factor 3, leading to significantly longer trace lengths and larger wire loading effects. Lack of a proper timing analysis during place and route led to setup/hold violations in shift registers.
- •Fixed version submitted using IBM "re-spin", modifying 3 mask layers to improve critical clock routing by hand. Six wafers from initial engineering run waiting for back-end processing will be processed with new mask set.
- •Initial wafer probing done, and yield is extremely high (92% for fairly complete set of tests, not yet including all analog functionality). This suggests (as have further technical investigations by IBM), that there are critical processing differences between the two foundries for our particular designs.

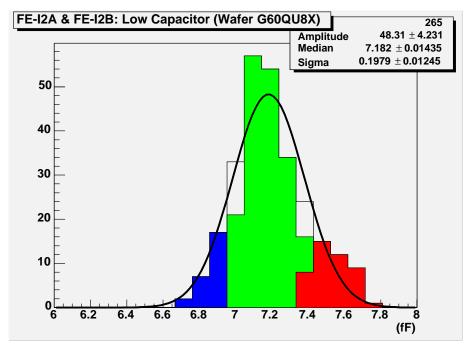
#### **Example probe results:**

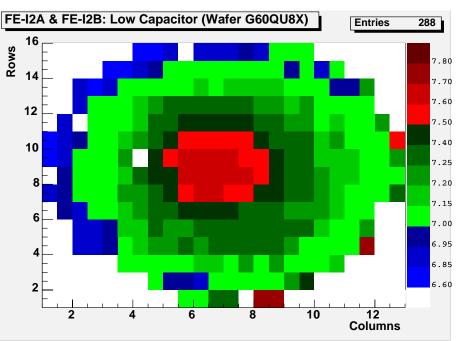
- Wafer layout includes 144 reticles with 288 potentially good die.
- After initial power tests, most critical tests involve validation of 231 bit Global Register, and 2880 bit Pixel Register with 40320 SEU-tolerant latches.



- •Pixel Register divided into 9 independent sub-units (column pairs). Negative value for test result indicates Global Register failed. Yield is above 94%, and problems are concentrated on extreme edge of wafer.
- Complete wafer probing includes extensive digital and analog tests, including digital hit injection to test readout with known data, characterization of all internal DACs, threshold/noise scans and TOT measurements. Takes about 20 hours/wafer.

• Implemented capacitance measurement circuit to measure critical metal capacitors in input pad (Clo, Chi, Cfb). for each chip and normalize charge injection:





- •This shows minute oxide thickness variations over wafers (a few percent in average thickness would explain the pattern seen). This is the ONLY test we have done so far which shows any systematic pattern over the wafer. Nice job Burlington !!!
- •Example here is small injection capacitor, which has value 7.2fF with an RMS of 0.2fF. Note this has increased from FE-I1 to a target value of 8fF.

## **Summary**

- •Lengthy design program has led to very sophisticated and high performance pixel arrays meeting all ATLAS pixel detector requirements. Final chips are implemented in commercial 0.25μ technology, using radiation-tolerant layout techniques to achieve 50MRad tolerance and good SEU hardness without latch-up or other fatal effects.
- •First FE-I1 DSM prototypes now extensively evaluated and essentially meet all ATLAS requirements. Evaluation process has included construction of about 60 pre-production modules, followed by lab measurements, testbeam measurements, and proton irradiations to full lifetime dose.
- •Present FE-I2 design is a second generation almost-production chip. It contains 3.5M transistors for 2880 channels of readout. Expect that the weaknesses observed in the first generation chip have been largely addressed, however chip has only been characterized on probe station up to now.
- •First experience with yield of new design looks excellent. After ten years of R&D, we are close to serial production for this project.